

EXHIBIT B

BEST AVAILABLE COPY

Jonathan P. L. Cox[○],
Ian M. Tomlinson^{○■△} and
Greg Winter^{○■}

MRC Centre for Protein
Engineering[○] and MRC Laboratory
of Molecular Biology[■], Cambridge

A directory of human germ-line V_{κ} segments reveals a strong bias in their usage

From the genomic DNA of a single individual, we have amplified, cloned and sequenced 37 human germ-line V_{κ} segments. Four of these segments were new. We then compiled a comprehensive directory of all germ-line V_{κ} segments and identified 50 different sequences with open reading frames. Comparison with 236 rearranged sequences revealed that no more than 24 of these germ-line sequences could be assigned rearranged counterparts, that some of these were rarely used, and that only about 11 sequences are used frequently. This suggests that the expressed V_{κ} repertoire is mainly derived from a limited number of segments. Most surprisingly, the J_{κ} -distal region of the locus appears to be rarely used: we could unambiguously assign 162 rearranged sequences to V_{κ} segments of the J_{κ} -proximal region, but only 5 to segments of the J_{κ} -distal region.

1 Introduction

Recently antibodies of predefined binding specificity have been derived from repertoires of associated heavy and light chain variable (V) domains displayed on the surface of filamentous phage ([1]; for review see [2]). The process mimics production of antibodies by the immune system and bypasses hybridoma technology and immunization [3, 4]. Diverse repertoires of V domains have been provided by PCR amplification [5] of heavy and light chain V genes [6] from populations of lymphocytes. Alternatively the repertoires have been provided by *in vitro* rearrangement of cloned V segments [7]. To provide a bank of cloned V gene segments, we have amplified and sequenced V_H and V_{κ} segments from genomic DNA, leading to the isolation of almost all the functional human V_H segments [8] and many of the human V_{κ} segments [9]. Here we have attempted to gather a large selection of human germ-line V_{κ} segments using the same strategy. After submission of this report, a number of new V_{κ} segments were sequenced and mapped [10, 11]. For completeness we have included this information in our revised paper.

The human immunoglobulin light chain kappa (κ) locus is located on the short arm of chromosome 2 (2p11-12) [12, 13] and consists of a C_{κ} gene, 5 J_{κ} segments and at least 76 mapped germ-line V_{κ} segments [11, 14]. The V_{κ} segments may be divided by sequence homology [15] into three main subgroups, I–III, and several smaller subgroups (IV, V, VI and VII) [11, 16–18]. The portion of the locus harboring the majority of V_{κ} segments is thought to have arisen from a duplication event [19, 20]: 36 segments are located in the J_{κ} -distal region and 40 in the J_{κ} -proximal region [11, 14, 21, 22]. Segments are clustered in four distinct regions, A, B, L

and O (Fig. 1) [18]. The A, L and O clusters are found within both the J_{κ} -distal and the J_{κ} -proximal regions, the B cluster only within the J_{κ} -proximal region.

Of the 76 mapped V_{κ} segments within the major locus, 57 have published sequences. Of these 57 segments, 48 have alleles with open reading frames (these correspond to 50 different sequences due to there being some segments from the J_{κ} -proximal and J_{κ} -distal regions which are identical and other segments which have multiple alleles) and 9 have frame shifts, stop codons or incomplete exons and are, therefore, regarded as pseudogenes. Most of the remaining 19 segments are also known to be pseudogenes [10, 16, 17, 19, 20, 23–36] (Fig. 1). There are also a number of orphon V_{κ} segments which have been mapped outside the major κ locus: W1–W11 [37, 38], V108 [39], Chr22-1 to Chr22-5 [40, 41], Chr1 [41], cos 118 [41] and Z1–Z4 [42, 43]. There are two further segments with open reading frames that have not been mapped (LFVK5 and LFVK431; L. Foroni, personal communication).

To clone the functional V_{κ} segments, we assembled a database of published germ-line sequences and then designed PCR primers to amplify each of the different subgroups. A germ-line V_{κ} segment consists (5' to 3') of a leader sequence (L), a leader intron, a continuation of the leader sequence (L') and the V_{κ} exon (Fig. 2). Within the exon there are three framework regions (FR) and three complementarity-determining regions (CDR). Adjoining the 3' end of the exon is a highly conserved heptamer recombination signal and a less conserved nonamer region, the two separated by 11–14 nucleotides. Forward PCR primers were based in the heptamer region and back primers on the overlap between the leader intron and L' region (Fig. 2 and Table 1). Using these primers we amplified, cloned and sequenced germ-line V_{κ} segments from the genomic DNA extracted from the peripheral white blood cells of a single donor (DP).

[I 12076]

[△] Supported by the Medical Research Council Human Genome Mapping Project.

Correspondence: I. M. Tomlinson, MRC Centre for Protein Engineering, Hills Road, Cambridge CB2 2QH, GB (Fax: 02 23 40 21 40)

Abbreviations: CDR: Complementarity-determining region. FR: Framework region

Key words: Human immunoglobulin kappa (V_{κ}) segments

2 Materials and methods

2.1 Primer design

Forward and back primers were designed for the V_{κ} I, II, III, IV and VI subgroups from alignments of published germ-line sequences. For the V_{κ} I subgroup, three cluster-specific

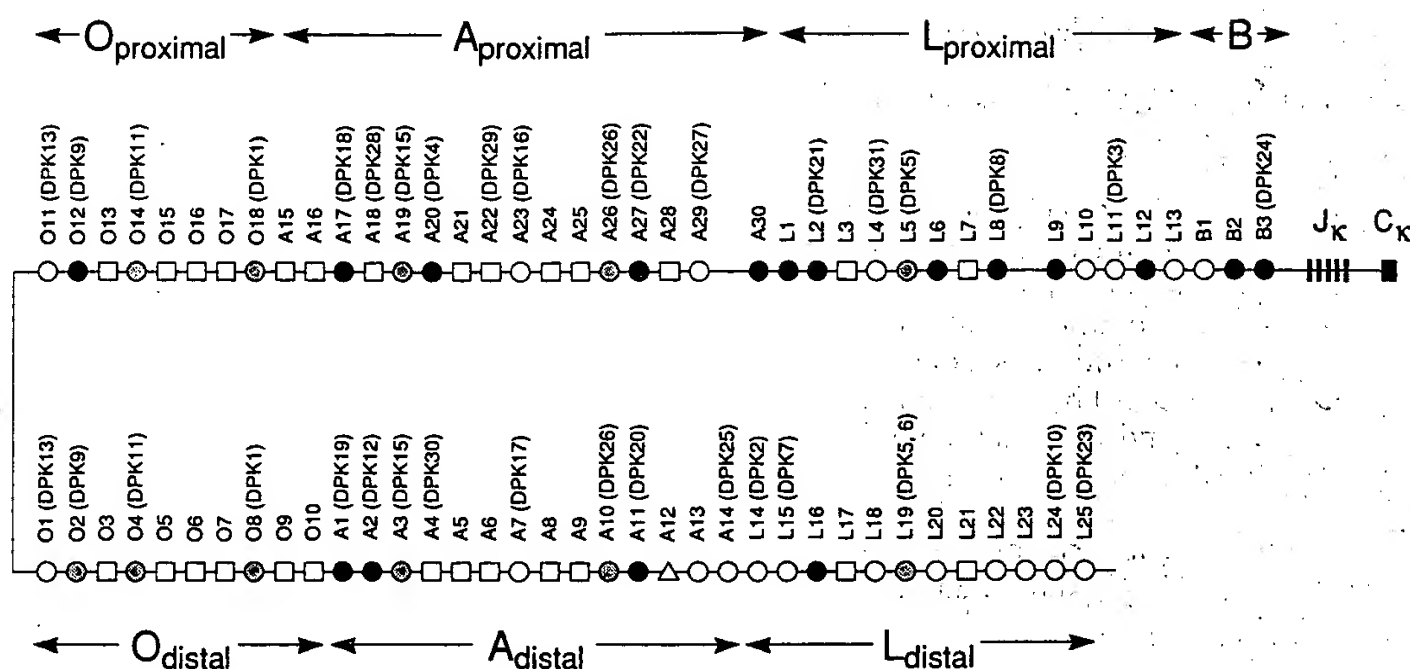


Figure 1. Schematic representation of the human immunoglobulin κ locus (after [11, 21]). Germ-line V_{κ} segments with open reading frames are represented by circles. The segments A4, A5, A18, A22, A28, L3, L7, L17 and L21 have frame shifts, stop codons or incomplete V_{κ} exons and are, therefore, regarded as pseudogenes (squares). The remaining 19 segments have either not been sequenced (triangle) or are known to be pseudogenes (squares). The segments L4 and L16 have a number of alleles, some with open reading frames and others with stop codons. Where a segment has been seen rearranged *in vivo*, the circle has been filled (black or shaded). If the rearranged sequences could not be unambiguously assigned to segments of the J_{κ} -proximal or J_{κ} -distal regions, the corresponding segments from both regions have been shaded. Sequences from this study have been attributed their DPK codes (see text).

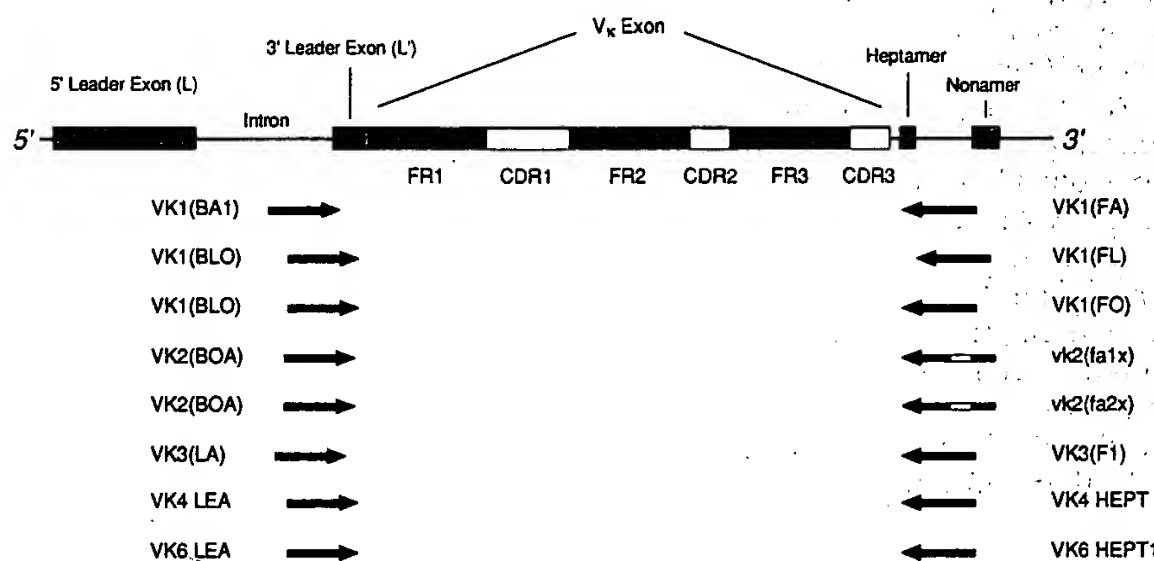


Figure 2. Location of PCR primers for amplification of the human germ-line V_{κ} segment.

Table 1. Primers used for PCR amplification of the V_{κ} exon

Primer	Sequence (5' to 3')
$V_{\kappa}I$ VK1(FA)	GCT CTA GAC GGG CTT GTA TCA CAG TG
VK1(FL)	GCT CTA GAG TT(CT) (AG)GG T(GT)(GT) GTA ACA CT
VK1(FO)	GCT CTA GAA TG(AC) CTT GT(TA) ACA CTG TG
VK1(BA1)	CCC CCA AGC TTT GTT CCT AAT ATC AGA TA
VK1(BLO)	CCC CCA AGC TTA ATC (TG)CA GGT (GT)CC AGA TG
$V_{\kappa}II$ vk2(fa1x)	GAG GTT TTC TAG A(TG)G (GA)(GT)(CT) TGT A(GC)C ACT GTG
vk2(fa2x)	GAG GTT TTC TAG AAG (GA)(GT)(CT) TGT A(GC)C ACT GTG
VK2(BOA)	CCC CCA AGC TT(TA) A(TC)T TCA GGA TCC AGT G
$V_{\kappa}III$ VK3(F1)	GGA ATT CT(CT) A(TA)G (CT)TG AAT CAC TGT G
VK3(LA)	CCC CCA AGC TTT CCA AT(TC) T(CT)(AG) GAT ACC AC
$V_{\kappa}IV$ VK4 HEPT	GCT CTA GAC GAG GCT GAA GCA CTG TG
VK4 LEA	CCC CCA AGC TTA CTA CAG GTG CCT ACG GG
$V_{\kappa}VI$ VK6 HEPT1	GCT CTA GAG GGT TGT A(GA)C ACA GTG TG
VK6 LEA	CCC CCA AGC TTT TTT CAG CCT CCA GGG GT

forward primers were designed, VK1(FA), VK1(FL), and VK1(FO), corresponding to the A, L and O clusters, respectively. Two cluster-specific back primers, VK1(BA1) and VK(BLO), corresponding to the A cluster and the combined B and L clusters were also used. *Hind*III restriction sites were incorporated into the 5' end of all back primers. All forward primers had *Xba*I restriction sites at the 5' end, except VK3(F1) which had an *Eco*RI site and the V_κII primers, where the *Xba*I site was introduced into the middle of the primer to replace a particularly degenerate section. The primers were used in pairs as indicated in Fig. 2.

2.2 PCR amplification and sequencing

PCR amplification was performed according to [8] with the following modifications. Products from the PCR amplifications and restriction enzyme digests were purified using MagicTM PCR Preps (Promega, Madison, WI). Inserts from plaques picked from a TYE [44] plate were amplified with an M13mp19-specific primer pair using the PCR (25 cycles, each cycle consisting of 1 min at 94 °C, 1 min at 55 °C, 30 s at 72 °C; at the end of 25 cycles there was a final extension at

65 °C for 5 min). Sequencing was performed using an M13-specific primer, Taq polymerase and fluorescent dideoxy chain terminators [45]. The sequences were analyzed on an Applied Biosystems 373A Automated DNA Sequencer (Foster City, CA).

3 Results

Amplification using the subgroup-specific primers with suitable adjustment of the annealing temperature gave single PCR bands of varying intensities. Primer pairs were found to be specific for their respective subgroups. We sequenced 142 clones (62 V_κI, 60 V_κII, 14 V_κIII, 1 V_κIV and 5 V_κVI). From these, 37 V_κ sequences were identified. Excluding those identical to mapped orthon segments (see below) 27 segments had open reading frames (DPK1-DPK27): 24 were identical to published sequences and 3 segments, DPK2, DPK14 and DPK23, were new. DPK2 was most similar to L1/HK137 (14 nucleotide changes), DPK14 to the pseudogene A5 (insertion of one nucleotide and 3 changes), and DPK23 to L10/Vh (8 nucleotide changes). The remaining segments are identical to the pseudogenes

Table 2. Assignment of rearranged V_κ sequences to their closest germline counterparts

Germline sequence	No. of rearranged sequences ^a	Closest rearranged sequence Δ(N, P) ^b
V_κI		
O8/O18/DPK1	21	(0, 0)
A30	3	(5, 2)
L1/HK137	1	(15, 8)
A20/DPK4	9	(0, 0)
L5/L19 ¹ /Vb/Vb'/V4b/DPK5	4	(3, 0)
L8/Vd/DPK8	10	(1, 0)
L9/Ve	3	(0, 0)
L12 ¹ /HK102/V1	1	(37, 17)
L12 ²	12	(5, 1)
O12 ¹ /V3b	1	(2, 1)
O2/O12 ² /DPK9	30	(0, 0)
O4/O14/LFVK19H/DPK11	1	(11, 6)
	<u>96</u>	
V_κII		
A2/DPK12	1	(9, 8)
A3/A19/DPK15	11	(0, 0)
A17/DPK18	11	(1, 0)
A1/DPK19	1	(0, 0)
	<u>24</u>	
V_κIII		
A11/humkv305/DPK20	1	(3, 2)
L16/humkv328/humkv328h2	2	(3, 3)
L2/humkv328h5/DPK21	20	(1, 0)
A27/humkv325/V _κ RF/DPK22	48	(0, 0)
L6/Vg	19	(0, 0)
	<u>90</u>	
V_κIV		
B3/VkIV/DPK24	22	(1, 1)
V_κV		
B2/EV15	2	(0, 0)
V_κVI		
A10/A26/DPK26	2	(3, 1)

a) Number of rearranged sequences assigned to their closest germline counterpart.

b) Δ(N, P) denotes the number of nucleotide (N) and amino acid residue changes (P).

				L1		L2		L3		
				FR1	CDR1	FR2	CDR2	FR3	CDR3	
				11020	30	40	50	607080	90	
V _H I	7	RO	08/018[33]/DPK1	DIQMTQSPSSLSASVGDRTITC	QASQDISN	YLN	WYQKPGKAPKLLIY	DASNLET	GVPSRFSGSGSGIDFTLTISLQPEDFATYYC	QQYDNL
	7		114[10]/DPK2	NIQMTQSPSAMSASVGDRTITC	RARQGISN	YLA	WYQKPGKVPKLLIY	AASSLQS	GVPSRFSGSGSGTEFTLTISLQPEDFATYYC	LQHNYP
	7		115 ¹ [10]/HK101[31]/HK146/HK189[19]	DIQMTQSPSSLSASVGDRTITC	RARQGISS	WLA	WYQKPGKAPKSLIY	AASSLQS	GVPSRFSGSGSGIDFTLTISLQPEDFATYYC	QQYNSYP
	7		111/VF[33]/DPK3	AIQMTQSPSSLSASVGDRTITC	RASQGIRN	DLG	WYQKPGKAPKLLIY	AASSLQS	GVPSRFSGSGSGIDFTLTISLQPEDFATYYC	LQDYNYP
	7	R2	A30[10]	DIQMTQSPSSLSASVGDRTITC	RASQGIRN	DLG	WYQKPGKAPKRLIY	AASSLQS	GVPSRFSGSGSGTEFTLTISLQPEDFATYYC	LQHNYP
	7		LFVK5[*]	DIQMTQSPSSLSASVGDRTITC	RASQGISN	NLN	WYQKPGKTPKLLIY	AAPSLQS	GIPSRFSDSGSGADYTLTISRLOPEDFATYYC	QQSDSTP
	7		LFVK431[*]	DIQMTQSPSSLSASVGDRTITC	RASQGISN	YLA	WYQKPGKAPKSLIY	AASSLQS	GVPKFSFGSGSGIDFTLTISRLOPEDFATYYC	QQYNSYP
	7	R8	11[10]/HK137[19]	DIQMTQSPSSLSASVGDRTITC	RASQGISN	YLA	WYQKPGKAPKSLIY	AASSLQS	GVPSRFSGSGSGIDFTLTISRLOPEDFATYYC	QQYNSYP
	7	RO	A20[17]/DPK4	DIQMTQSPSSLSASVGDRTITC	RASQGISN	YLA	WYQKPGKVPKLLIY	AASTLQS	GVPSRFSGSGSGIDFTLTISRLOPEDFATYYC	QQYNSAP
	7		118 ¹ [10]/Va ¹ [20]	AIQLTQSPSSLSASVGDRTITC	RASQGISS	ALA	WYQKPGKAPKLLIY	DASSLES	GVPSRFSGSGSGIDFTLTISRLOPEDFATYYC	QQFNYP
	7		14/118 ² /Va ¹ /V4a[35]	AIQLTQSPSSLSASVGDRTITC	RASQGISS	ALA	WYQKPGKAPKLLIY	DASSLES	GVPSRFSGSGSGIDFTLTISRLOPEDFATYYC	QQFNYP
	7	RO	15/119 ¹ [10]/Vb/Vb ¹ /V4b[35]/DPK5	DIQMTQSPSSVSASVGDRTITC	RASQGISS	WLA	WYQKPGKAPKLLIY	AASSLQS	GVPSRFSGSGSGIDFTLTISRLOPEDFATYYC	QQANSFP
	7		119 ² /Vb ¹ [35]/DPK6	DIQMTQSPSSVSASVGDRTITC	RASQGISS	WLA	WYQKPGKAPKLLIY	AASSLQS	GVPSRFSGSGSGIDFTLTISRLOPEDFATYYC	QQANSFP
	7		115 ² [10]/HK134/HK166[19]/DPK7	DIQMTQSPSSLSASVGDRTITC	RASQGISS	WLA	WYQKPGKAPKSLIY	AASSLQS	GVPSRFSGSGSGIDFTLTISRLOPEDFATYYC	QQYNSYP
	7	RO	18/Vd[35]/DPK8	DIQLTQSPSFLSASVGDRTITC	RASQGISS	YLA	WYQKPGKAPKLLIY	AASTLQS	GVPSRFSGSGSGTEFTLTISRLOPEDFATYYC	QQYNSYP
	7	RO	19/Ve[35]	AIRMTPSPSSSASTGDRVITTC	RASQGISS	YLA	WYQKPGKAPKLLIY	AASTLQS	GVPSRFSGSGSGIDFTLTISRLOPEDFATYYC	QQYNSYP
	7	R17	112 ¹ /HK102[31]/V1[25]	DIQMTQSPSTLSASVGDRTITC	RASQSISS	WLA	WYQKPGKAPKLLIY	DASSLES	GVPSRFSGSGSGTEFTLTISRLOPEDFATYYC	QQYNSYS
	7	R1	112 ² [10]	DIQMTQSPSTLSASVGDRTITC	RASQSISS	WLA	WYQKPGKAPKLLIY	KASSLES	GVPSRFSGSGSGTEFTLTISRLOPEDFATYYC	QQYNSYS
7	R1	012 ¹ /V3b[29]	DIQMTQSPSFLSASVGDRTITC	RASQSISS	YLN	WYQKPGKAPKLLIY	AASSLQS	GVPSRFSGSGSGIDFTLTISRLOPEDFATYYC	QQGYSTP	
7	RO	02/012 ² [29]/DPK9	DIQMTQSPSSLSASVGDRTITC	RASQSISS	YLN	WYQKPGKAPKLLIY	AASSLQS	GVPSRFSGSGSGIDFTLTISRLOPEDFATYYC	QQSYSTP	
7		124[10]/Ve ¹ /V13[25]/DPK10	VIVMTQSPSLLSASTGDRVITTC	RMSQGISS	YLA	WYQKPGKAPPELLIY	AASTLQS	GVPSRFSGSGSGIDFTLTISRLOPEDFATYYC	QQYNSFP	
7	R6	04/014[29]/LFVK19H[*]/DPK11	DIQLTQSPSSLSASVGDRTITC	RVSQGISS	YLN	WYQKPGKVPKLLIY	SASNLQS	GVPSRFSGSGSGIDFTLTISRLOPEDFATYYC	QQYNSAP	
7		122[10]	DIQMTQSPSFLSASVGDRTITC	WASEGISS	NLA	WYQKPGKSPKLLIY	DAKDLHP	GVSRRFSGRSGSGIDFTLTISRLOPEDFATYYC	QQDFSTP	
7		123[10]	AIRMTPSPSFLSASVGDRTITC	WASQGISS	YLA	WYQKPGKAPKLLIY	YASSLQS	GVPSRFSGSGSGIDFTLTISRLOPEDFATYYC	QQYNSYP	
V _H II	12	R8	A2[23]/DPK12	DIVMTQTPSLSPVTPGPASISC	KSSQSLHS	DGTIYLY	WYQKPGQPPQLLIY	EVSNRFS	GVPDRFSGSGSGIDFTLTISRVEAEDVGYYC	MQSTQLP
	13		01/011 ¹ [29]/DPK13	DIVMTQTPSLSPVTPGPASISC	RSSQSLHS	DGTIYLY	WYQKPGQSPQLLIY	TLSTRAS	GVPDRFSGSGSGIDFTLTISRVEAEDVGYYC	MQRIEFP
	13		011 ² /V3a[29]	ASISC	RSSQSLHS	DGTIYLY	CYQKPGQSPQLLIY	TLSTRAS	GVPDRFSDSGSGIDFTLTISRVEAEDVGYYC	MQRIEFP
	12		113[10]	DIVMTQHLLSLPIPLGPASISC	RSSQSLHS	DGTIYLY	WYQKPGQSPQLLIY	TTSNKFY	GVPNKFSGSRSGTGFTLKFSKVEAEDVGYYC	EQGLQGP
	12		DPK14	EIVMTQTPSLSPVTPGPASISC	RSSQSLHS	DGTIYLY	WYQKPGQPPQLLIY	EVSNRFS	GVPDRFSGSGSGIDFTLTISRVEAEDVGYYC	MQDAQDP
	12	RO	A3[24]/A19[17]/DPK15	DIVMTQSPSLSPVTPGPASISC	RSSQSLHS	NGYNYLY	WYQKPGQSPQLLIY	LGSNRAS	GVPDRFSGSGSGIDFTLTISRVEAEDVGYYC	MQALQTP
	12		A29[17]/DPK27	DIVMTQTPSPVTPGPASISC	RSSQSLHS	NGYTYLY	WYQKPGQSPQLLIY	RVSNHLS	GVPDRFSGSGSGIDFTLTISRVEAEDVGYYC	MQATQFP
	12		A13[17]	DIVMTQTPSPVTPGPASISC	RSSQSLHS	NGYTYLY	WYQKPGQSPQLLIY	RVSNRFS	GVPDRFSGSGSGIDFTLTISRVEAEDVGYYC	MQATQFP
	12		A23[24]/DPK16	DIVMTQTPSLSPVTPGPASISC	RSSQSLHS	DGTIYLY	WYQKPGQPPQLLIY	KISNRFS	GVPDRFSGSGAGIDFTLTISRVEAEDVGYYC	MQATQFP
	12		A7[17]/DPK17	DIVMTQTPSLSPVTPGPASISC	RSSQSLHS	DGTIYLY	WYQKPGQPPQLLIY	KVSNRFS	GVPDRFSGSGAGIDFTLTISRVEAEDVGYYC	MQATQFP
	12	RO	A17[17]/DPK18	DIVMTQSPSLSPVTPGPASISC	RSSQSLVYS	DGTIYLY	WYQKPGQSPRLLIY	KVSNRFS	GVPDRFSGSGSGIDFTLTISRVEAEDVGYYC	MQGTHWP
	12	RO	A1[17]/DPK19	DIVMTQSPSLSPVTPGPASISC	RSSQSLVYS	DGTIYLY	WYQKPGQSPRLLIY	KVSNRFS	GVPDRFSGSGSGIDFTLTISRVEAEDVGYYC	MQGTHWP
V _H III	8	R2	A11[24]/humkv305[27]/DPK20	EIVLTQSPATLSLSPGERATLSC	GASQSVSS	YLA	WYQKPGQAPRLLIY	DASSRAT	GIPDRFSGSGSGIDFTLTISRLEPEDFATYYC	QQYSSSP
	7		120/Vg ¹ [20]	EIVLTQSPATLSLSPGERATLSC	RASQGVSS	YLA	WYQKPGQAPRLLIY	DASNAT	GIPARFSGSGPGIDFTLTISRLEPEDFATYYC	QQRSNWH
	7	R3	116[10]/humkv328/humkv328h2[34]	EIVMTQSPATLSLSPGERATLSC	RASQSVSS	NLA	WYQKPGQAPRLLIY	GASTRAT	GIPARFSGSGSGTEFTLTISRLOPEDFATYYC	QQYNNWP
	7	RO	12[10]/humkv328h5[34]/DPK21	EIVMTQSPATLSLSPGERATLSC	RASQSVSS	NLA	WYQKPGQAPRLLIY	GASTRAT	GIPARFSGSGSGTEFTLTISRLOPEDFATYYC	QQYNNWP
	8	RO	A27[24]/humkv325/Vh RF[28]/DPK22	EIVLTQSPGTLSPGERATLSC	RASQSVSS	YLA	WYQKPGQAPRLLIY	GASSRAT	GIPDRFSGSGSGIDFTLTISRLEPEDFATYYC	QQYSSSP
	8		125[10]/DPK23	EIVMTQSPATLSLSPGERATLSC	RASQSVSS	YLS	WYQKPGQAPRLLIY	GASTRAT	GIPARFSGSGSGIDFTLTISRLOPEDFATYYC	QQDYNLP
	8		110 ¹ [10]	EIVMTQSPPTLSLSPGERATLSC	RASQSVSS	YLS	WYQKPGQAPRLLIY	GASTRAT	GIPARFSGSGSGIDFTLTISRLOPEDFATYYC	QQDYNLP
	8		110 ² /Vh[32]	EIVMTQSPPTLSLSPGERATLSC	RASQSVSS	YLT	WYQKPGQAPRLLIY	GASTRAT	SIPARFSGSGSGIDFTLTISRLOPEDFATYYC	QQDYNLP
7	RO	16/Vg[32]	EIVLTQSPATLSLSPGERATLSC	RASQSVSS	YLA	WYQKPGQAPRLLIY	DASNAT	GIPARFSGSGSGIDFTLTISRLEPEDFATYYC	QQRSNWP	
V _H IV	13	R1	B3/VhIV[16]/DPK24	DIVMTQSPDSLAVSLGERATLSC	KSSQSVLYSSNNKLYLA	WYQKPGQPPQLLIY	WASTRES	GVPDRFSGSGSGIDFTLTISRLOAEDVAVYYC	QQYNSYP	
V _H V	7	RO	B2/EV15[30]	ETTLTQSPAFMSATPGDKVNISC	KASQIDDD	DMN	WYQKPGQAPRLLIY	EATTLVP	GIPPRFSGSGYGTIDFTLTISRLEPEDFATYYC	LQHDNFP
V _H VI	7		A14[26]/DPK25	DIVMTQSPAFSLSPVTPGERATLSC	QASEGIN	YLY	WYQKPGQAPRLLIY	YASQSI	GVPSRFSGSGSGIDFTLTISRLEAEDATYYC	QQGNKHP
	7	R1	A10[26]/A26[17]/DPK26	EIVLTQSPDFQSVTPKEKVTITC	RASQSIS	SLH	WYQKPGQSPKLLIY	YASQSF	GVPSRFSGSGSGIDFTLTISRLEAEDATYYC	HQSSSLP
V _H VII	11		B1[36]	DIVLTQSPASLAVSPQGRATITC	RASESVSFL	GINLIH	WYQKPGQPPQLLIY	QASNKOT	GVPARFSGSGSGIDFTLTISRNPVANDTANYC	LQSKNFP

Figure 3. Amino acid sequences of germ-line V_H segments with open reading frames. Where possible, sequences have been assigned to their respective loci (bold type). Where two alleles have different nucleotide sequences these are followed by the superscripts ¹ or ². Amino acid sequences from this study are labeled DPK1-DPK27 (see text). References for other sequences are in brackets, [*] indicates L. Foroni, personal communication. Amino acids are shown in single-letter codes. Sequences are arranged alphabetically within each subgroup according to the sequence of CDR1. Where two CDR1 sequences are identical, the order is based similarly upon the sequence of CDR2 or CDR3. The hypervariable regions, CDR1, CDR2 and CDR3 as defined by Kabat et al. [15] are labeled, as are the antigen-binding loops, L1, L2 and L3 as defined by Chothia and Lesk [50]. Numbering of amino acid residues is according to [15] except in CDR1 where numbering is according to [50]. The length of the L1 loop [50] is shown in italics. Germ-line sequences with known rearranged counterparts have been labeled RN, where N represents the number of amino acid differences between the germ-line V_H sequence and its closest rearranged counterpart (see Table 2). Note that L5/L19¹/Vb/Vb¹/V4b/DPK5 and L19²/Vb¹/DPK6; and L16/humkv328/humkv328h2 and L2/humkv328h5/DPK21 have identical protein sequences but different nucleotide sequences (see Fig. 4).

Figure 4

Figure 4

[illegible][illegible]

	+1	+10	+20	CHRI	+40
<i>V_hIII</i>					
AAA [24] / humkv305 [27] / DEK20	E I V L T Q S P A T L S L S P G E R A T L S C G A S Q S V S S Y L A W Y Q Q K P G L A P R L L I Y	GAA ATT GTG TTG ACG CAG TCT CCA GGC ACC CTC TCT TTG TCT CCA GGG GAA AGA GGC ACC CTC TCC TCC GGG GGC AGT CAG AGT GTT ACC AGC AGC TAC TTA GGC TGG TAC CAG CAG AAA CTT GGC CTG GCG CCG AGS CTC ATC TAT			
120/Vg* [20]
116 [10] / humkv328 / humkv328b.2 [34]
12 [10] / humkv328b.5 [34] / DEK21
127 [24] / humkv325/Vh 18 [28] / DEK22
125 [10] / DEK23
110 ¹ [10]
110 ² / Vh [32]
16/Vg [32]

[illegible]

[illegible]

Figure 4. Nucleotide sequences of germ-line V_{α} segments with open reading frames. Sequences were extracted from the Genbank/EMBL nucleotide databases wherever possible. Sequences from this study have been labeled with their DPK codes (see text). References for other sequences are in brackets. [*] denotes L. Foroni, personal communication. A translation of the master sequence is shown directly above it: residues are numbered according to [15]. Nucleotides identical to the master sequence are represented by a dot and insertions are represented by a dash. The directory of human germ-line V_{α} sequences is available on disk on request.

A4 [25], A18 [19], A22 [24], L4 (Va) [35], HK100 [31] and to the orphon segments W2 [38], W8/W10 [38], Z2 [43], chr22-4 [41]: DPK37 is new and is most similar to the orphon Z3 [41] (deletion of one nucleotide and three changes). The sequences of the four new segments, DPK2, DPK14, DPK23 and DPK37, were confirmed with clones from independent PCR amplifications. Subsequently, two of these segments, DPK2 and DPK23, proved to be identical to the sequences L14 and L25, which were published after the submission of this paper [10]. The fact that 35 of the 37 segments amplified from DP are identical to published sequences confirms that germ-line V_α sequence polymorphism is rather limited [46].

In addition to the confirmed sequences, many clones had sequences which were not observed in independent amplifications. Some had one or two nucleotide differences from the V_α segments DPK1-DPK37 and were probably due to PCR errors. The 21 nucleotide substitutions in the 52 V_α I sequences correspond to 5×10^{-5} changes per nucleotide per PCR cycle and are consistent with the error rate for Taq polymerase [47]. Other clones appeared to be due to PCR cross-over [8, 48].

We have compiled a comprehensive directory of the amino acid and nucleotide sequences of all germ-line V_α segments with open reading frames (Figs. 1, 3 and 4). Where possible, sequences have been assigned to their respective loci. We have compared these sequences with those of 113 rearranged V_α sequences taken from the Genbank/EMBL nucleotide databases and 123 additional sequences from the literature (our rearranged database is available on request). All rearranged sequences were assigned to their closest germ-line counterparts. In Table 2 we note the closest sequence identities and the number of rearranged sequences assigned to each germ-line sequence. In a few examples, the rearranged sequence was found to be a composite of two V_α segments, presumably arising by PCR cross-over during cloning. For example, V_α I-X14 [49] is identical to O4/O14/LFVK19H/DPK11 over the first 220 nucleotides and virtually identical to O2/O12²/DPK9 over the last 100 nucleotides, due to a cross-over in FR3.

Only 24 of the 50 germ-line sequences in our directory were found to have rearranged counterparts. Although 162 rearranged sequences correspond to mapped V_α segments from the J_α -proximal region of the kappa locus only 5 are derived from the J_α -distal region. These correspond to the germ-line sequences A2/DPK12, A1/DPK19, A11/humkv-305/DPK20 and L16/humkv328/humkv328h2 (see Fig. 1). Of the rearranged sequences, 69 correspond to O8/O18/DPK1, L5/L19/Vb/Vb'/V4b/DPK5, O2/O12²/DPK9, O4/O14/LFVK19H, A3/A19/DPK15 and A10/A26/DPK26: in these cases each rearranged sequence could have been derived from V_α segments from either the J_α -proximal or J_α -distal region. No rearranged sequences correspond to the unmapped V_α segments LFVK5, LFVK431 and DPK14.

The distribution of the number of amino acid differences across all 236 assigned rearranged sequences is shown in Fig. 5. This includes sequences of antibodies with a wide range of specificities. As sequence polymorphism at the germ-line level is limited (see above) the majority of changes are probably due to somatic mutation.

4 Discussion

4.1 Functional V_α segments

The map of the V_α locus (Fig. 1) contains 76 segments. It is estimated, on the basis of hybridization experiments, that 4–6 V_α segments have yet to be mapped [11, 14]. The V_α segments LFVK5, LFVK431 and DPK14 could, therefore, correspond to the unmapped loci or could be allelic variants of mapped V_α segments.

From our sequence directory (Fig. 3 and 4) and from the assignment of 236 rearranged sequences to their closest germ-line counterparts (Table 2) we find that no more than 24 of the 50 germ-line sequences have rearranged counterparts, suggesting that the expressed V_α repertoire is mainly derived from these segments. Indeed for 7 of these V_α sequences we could find only a single example of a rearrangement, and in 4 cases (L1/HK137, L12¹/HK102/V1, O4/O14/LFVK19H/DPK11 and A2/DPK11) the assignment is tentative in view of the number of sequence differences. It is also possible that some of the remaining segments with open reading frames are functional, but are only rarely used. Others may well prove to have frame shifts or stop codons in the leader exon, defective recombination signals or splice sites or other *cis*-acting defects which render them non functional. A14/DPK25, for example, has an altered heptamer which may prevent its rearrangement. We have made similar observations for human germ-line V_H segments where only 49/83 germ-line sequences with open reading frames were seen as rearranged genes [8].

4.2 V_α segment usage

The assignment of rearranged V_α genes to their germ-line counterparts (Table 2) confirms that segments from all four clusters (A, B, L and O) and from six subgroups (I–VI) are used. Segments from the V_α I and V_α III subgroups are used most frequently (96 and 90 rearranged sequences respectively); those from the V_α II and V_α IV subgroups less frequently (24 and 22 respectively); and those from the V_α V and V_α VI subgroups rarely (2 rearranged sequences each). The only V_α VII segment, B1, has no rearranged counterpart in our database. As noted above, some of the V_α

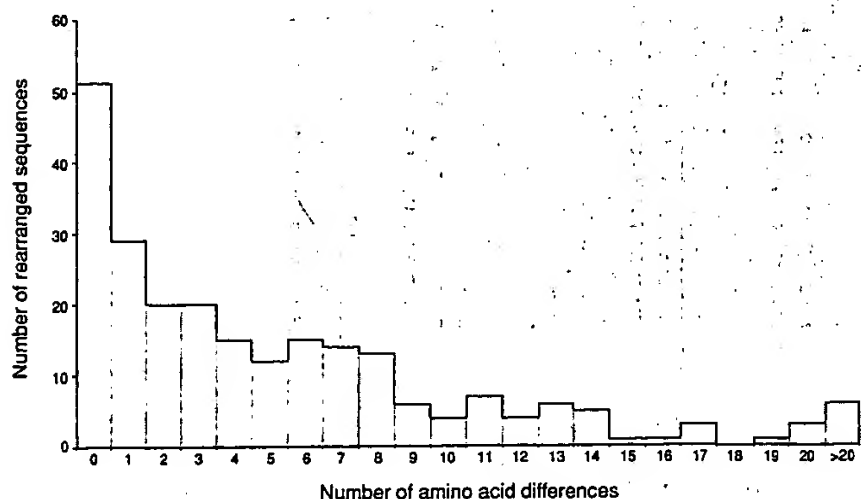


Figure 5. Distribution of the number of amino acid differences between each sequence in a database of rearranged V_α sequences (236 in total) and its closest germ-line counterpart in our directory. The database of rearranged V_α sequences is available on disk on request.

segments are rarely used, and of the 24 germ-line sequences, only about 11 appear to be frequently used (Table 2).

Of the rearranged sequences, 162 correspond to V_κ segments from the J_κ-proximal region of the kappa locus, whereas only 5 appear to be derived from the J_κ-distal region. Indeed, one of these rearranged sequences from the J_κ-distal region, EVJK11, which corresponds to A11/humkv305/DPK20, is the product of aberrant recombination [30]. An additional 69 rearranged sequences could conceivably be derived from V_κ segments in either the proximal or distal regions. However, as has been previously noted, the segments O2 (distal) and O12 (proximal) differ by one nucleotide in a region which flanks the exon [29]: rearranged sequences which include this region appear to be derived from O12 rather than O2 [29]. Although there is a clear bias towards use of the J_κ-proximal V_κ segments, we can detect no bias within the J_κ-proximal region.

The reason for the remarkable bias towards use of V_κ segments from the J_κ-proximal region is unclear. It may be due to the relative distances of the proximal and distal regions from the J_κ segments or the different recombination mechanisms necessary to produce the rearranged V_κ gene: most V_κ segments in the proximal region rearrange by deletion, whereas those in the distal region must rearrange by inversion. We note that lack of the J_κ-distal copy of the locus (haplotype 11) does not appear to be deleterious [21].

The assignment of rearranged V_κ genes may help in dissecting the mechanisms of the human immune system. For example, more than half of the rearranged sequences in our database have three or fewer residue changes (Fig. 5). Antibodies from patients with chronic lymphocytic leukemia (CLL) and X-linked agammaglobulinemia (XLA) are rarely mutated or unmutated. More highly mutated V_κ genes tend to be seen in myelomas and in antibodies which are subject to an antigen-driven response. In the case of autoantibodies, the level of somatic mutation appears to be normal.

4.3 Structures of loops implicit in the human V_κ segments

The antigen-binding loops of immunoglobulin variable domains have been shown to adopt a limited number of main chain conformations or "canonical structures" [50, 51]. The structure of each loop depends on its length and the identity of certain key residues involved in its packing, and using this information it is sometimes possible to predict the structure of the loops from the sequence of the V_κ domain [51]. Here we have attempted to identify the loop structures of those human V_κ segments that appear to undergo rearrangement (Table 2).

Across different species, the V_κ antigen binding site loop L1 (residues 26–32, corresponding to CDR1, see Fig. 3) can have lengths of 6, 7, 8, 11, 12 and 13 residues and presumably can form at least six major conformations [50, 51]. The structures of antibodies with loops of 6, 7, 12 and 13 residues have been solved crystallographically: in each case residue 29 in the loop is buried in the β-sheet framework, and packs against residues 2, 25, 33 and 71 [50, 51]. These

packing contacts are often very similar: Ile/Val 2, Ala/Ser 25, Val/Ile/Leu 29, Leu/Met 33 and Tyr/Phe 71 [50].

However, for those human V_κ segments with rearranged counterparts (Table 2), we only see L1 loops of 7, 8, 12 and 13 residues (Fig. 3). In all cases the packing contacts are highly conserved, suggesting that the segments should encode four major conformations of the L1 loop. Of the germ-line sequences, 17 (141 rearranged sequences) encode a 7-residue L1 loop, 2 sequences (49 rearranged sequences of which 48 use the V_κIII segment A27) encode an 8-residue loop, 4 sequences (24 rearranged sequences) encode a 12-residue loop and 1 sequence (22 rearranged sequences) encodes a 13-residue L1 loop. L1 lengths of 6 residues have only been seen in mice [15], suggesting that mouse V_κ segments encode structures which cannot be encoded by human V_κ segments.

The L2 loop (residues 50–52, corresponding to CDR2, see Fig. 3) is only three residues long in all structures, and undergoes packing interactions with framework residues 48 and 64 [50, 51]. All germ-line V_κ sequences with rearranged counterparts (Fig. 3 and Table 2) have Ile 48 and Gly 64, indicating that they are likely to encode a single loop conformation.

The L3 loop (residues 91–96, corresponding to CDR3, see Fig. 3) is encoded mainly by the V_κ segment, and is most often a six-residue loop with Gln, Asn or His at 90 and Pro at 95 [50, 51]. Almost all human germ-line V_κ sequences with rearranged counterparts encode Gln 90 and Pro 95, except A20/DPK4 (Lys 90), O4/O14/LFVK19H/DPK11 (Arg 90), L12¹/HK102/V1 and L12² (Ser 95). This indicates that most human V_κ germ-line segments encode a single conformation of this loop: but depending on the location of the V-J join and nucleotide addition, other conformations may be formed in the rearranged gene.

We note that several of the germ-line sequences that do not have rearranged counterparts in our database (Fig. 3 and Table 2) have atypical residues at the key residues involved in the packing of the antigen-binding loops (L1 loop: L22 [Leu 48]; LFVK5 and O11²/V3a [Asp 64]; L3 loop: L20/Vg¹ [His 95]) or unusual loop lengths (B1 has an L1 of 11 residues).

In combination, the three antigen-binding loops L1, L2 and L3 of human germ-line V_κ segments are likely to encode four major folds. Since our repertoire of cloned human germ-line V_κ segments includes 15 of the 24 germ-line V_κ sequences with rearranged counterparts and examples of each of the four major folds it should be a valuable resource for building synthetic antibodies for use in phage display libraries.

We would like to thank D. J. Perry for numerous donations of his blood, M. Llewelyn, C. Brown, G. Walter, P. Dear and G. Cook for their help. We would like to thank H. G. Zachau for providing L-region sequence data after first submission of our paper and L. Foroni for unpublished sequences. We would also like to thank C. Chothia for his comments. EMBL Data Library accession numbers for the sequences DPK2, DPK14, DPK23 and DPK37 are Z27498–Z27501.

Received July 14, 1993; in final revised form December 21, 1993; accepted December 23, 1993.

5 References

- 1 McCafferty, J., Griffiths, A. D., Winter, G. and Chiswell, D. J., *Nature* 1990. 348: 552.
- 2 Hoogenboom, H. R., Marks, J. D., Griffiths, A. D. and Winter, G., *Immunol. Rev.* 1992. 130: 41.
- 3 Clackson, T., Hoogenboom, H. R., Griffiths, A. D. and Winter, G., *Nature* 1991. 352: 624.
- 4 Marks, J. D., Hoogenboom, H. R., Bonnert, T. P., McCafferty, J., Griffiths, A. D. and Winter, G., *J. Mol. Biol.* 1991. 222: 581.
- 5 Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. and Arnheim, N., *Science* 1985. 230: 1350.
- 6 Orlandi, R., Gussow, D. H., Jones, P. T. and Winter, G., *Proc. Natl. Acad. Sci. USA* 1989. 86: 3833.
- 7 Hoogenboom, H. R. and Winter, G., *J. Mol. Biol.* 1992. 227: 381.
- 8 Tomlinson, I. M., Walter, G., Marks, J. D., Llewelyn, M. B. and Winter, G., *J. Mol. Biol.* 1992. 227: 776.
- 9 Williams, S. C. and Winter, G., *Eur. J. Immunol.* 1993. 23: 1456.
- 10 Huber, C., Schäble, H. F., Huber, E., Klein, R., Meindl, A., Thiebe, R., Lamm, R. and Zachau, H. G., *Eur. J. Immunol.* 1993. 23: 2868.
- 11 Huber, C., Huber, E., Lautner-Rieske, A., Schäble, H. F. and Zachau, H. G., *Eur. J. Immunol.* 1993. 23: 2860.
- 12 Malcolm, S., Barton, P., Murphy, C., Ferguson-Smith, M. A., Bentley, D. L. and Rabbitts, T. H., *Proc. Natl. Acad. Sci. USA* 1982. 79: 4957.
- 13 McBride, O. W., Hieter, P. A., Hollis, G. F., Swan, D., Otey, M. C. and Leder, P., *J. Exp. Med.* 1982. 155: 1480.
- 14 Meindl, A., Klobeck, H.-G., Ohnheiser, R. and Zachau, H. G., *Eur. J. Immunol.* 1990. 20: 1855.
- 15 Kabat, E. A., Wu, T. T., Perry, H. M., Gottesman, K. S. and Foeller, C., *Sequences of Proteins of Immunological Interest*, U.S. Department of Health and Human Services, NIH, USA 1991.
- 16 Klobeck, H.-G., Bornkamm, G. W., Combriato, G., Mocikat, R., Pohlenz, H.-D. and Zachau, H. G., *Nucleic Acids Res.* 1985. 13: 6515.
- 17 Lautner-Rieske, A., Huber, C., Meindl, A., Pargent, W., Schäble, K. F., Thiebe, R., Zocher, I. and Zachau, H. G., *Eur. J. Immunol.* 1992. 22: 1023.
- 18 Zachau, H. G., *The Immunoglobulin Genes*, Academic Press, London and San Diego 1989, p. 91.
- 19 Bentley, D. L. and Rabbitts, T. H., *Cell* 1983. 32: 181.
- 20 Pech, M., Smola, H., Pohlenz, H.-D., Straubinger, B., Gerl, R. and Zachau, H. G., *J. Mol. Biol.* 1985. 183: 291.
- 21 Pargent, W., Schäble, K. F. and Zachau, H. G., *Eur. J. Immunol.* 1991. 21: 1829.
- 22 Zachau, H. G., *Biol. Chem. Hoppe Seyler* 1990. 371: 1.
- 23 Scott, M. G., Crimmins, D. L., McCourt, D. W., Zocher, I., Thiebe, R., Zachau, H. G. and Nahm, M. H., *J. Immunol.* 1989. 143: 4110.
- 24 Straubinger, B., Huber, E., Lorenz, W., Osterholzer, E., Pargent, W., Pech, M., Pohlenz, H.-D., Zimmer, F.-J. and Zachau, H. G., *J. Mol. Biol.* 1988. 199: 23.
- 25 Jaenichen, H.-R., Pech, M., Lindenmaier, W., Wildgruber, N. and Zachau, H. G., *Nucleic Acids Res.* 1984. 12: 5249.
- 26 Straubinger, B., Thiebe, R., Huber, C., Osterholzer, E. and Zachau, H. G., *Biol. Chem. Hoppe Seyler* 1988. 369: 601.
- 27 Chen, P. P., Albrandt, K., Orida, N. K., Radoux, V., Chen, E. Y., Schrantz, R., Liu, F. T. and Carson, D. A., *Proc. Natl. Acad. Sci. USA* 1986. 83: 8318.
- 28 Radoux, V., Chen, P. P., Sorge, J. A. and Carson, D. A., *J. Exp. Med.* 1986. 164: 2119.
- 29 Pargent, W., Meindl, A., Thiebe, R., Mitzel, S. and Zachau, H. G., *Eur. J. Immunol.* 1991. 21: 1821.
- 30 Stavnezer, J., Kekish, O., Batter, D., Grenier, J., Balazs, I., Henderson, E. and Zegers, B. J., *Nucleic Acids Res.* 1985. 13: 3495.
- 31 Bentley, D. L. and Rabbitts, T. H., *Nature* 1980. 288: 730.
- 32 Pech, M. and Zachau, H. G., *Nucleic Acids Res.* 1984. 12: 9229.
- 33 Scott, M. G., Crimmins, D. L., McCourt, D. W., Chung, G., Schäble, K. F., Thiebe, R., Quenzel, E.-M., Zachau, H. G. and Nahm, M. H., *J. Immunol.* 1991. 147: 4007.
- 34 Liu, M. F., Robbins, D. L., Crowley, J. J., Sinha, S., Kozin, F., Kipps, T. J., Carson, D. A. and Chen, P. P., *J. Immunol.* 1989. 142: 688.
- 35 Pech, M., Jaenichen, H.-R., Pohlenz, H.-D., Neumaier, P. S., Klobeck, H.-G. and Zachau, H. G., *J. Mol. Biol.* 1984. 176: 189.
- 36 Lorenz, W., Schäble, K. F., Thiebe, R., Stavnezer, J. and Zachau, H. G., *Mol. Immunol.* 1988. 25: 479.
- 37 Pohlenz, H.-D., Straubinger, B., Thiebe, R., Pech, M., Zimmer, F.-J. and Zachau, H. G., *J. Mol. Biol.* 1987. 193: 241.
- 38 Zimmer, F.-J., Huber, C., Quenzel, M., Schek, H., Stiller, C., Thiebe, R. and Zachau, H. G., *Biol. Chem. Hoppe Seyler* 1990. 371: 283.
- 39 Huber, C., Thiebe, R., Hameister, H., Smola, H., Lötscher, E. and Zachau, H. G., *Nucleic Acids Res.* 1990. 18: 3475.
- 40 Lötscher, E., Grzeschik, K.-H., Bauer, H. G., Pohlenz, H.-D., Straubinger, B. and Zachau, H. G., *Nature* 1986. 320: 456.
- 41 Lötscher, E., Zimmer, F.-J., Klopstock, T., Grzeschik, K.-H., Jaenichen, R., Straubinger, B. and Zachau, H. G., *Gene* 1988. 69: 215.
- 42 Straubinger, B., Thiebe, R., Pech, M. and Zachau, H. G., *Gene* 1988. 69: 209.
- 43 Rösenthaller, F., Schäble, K. F., Thiebe, R. and Zachau, H. G., *Biol. Chem. Hoppe Seyler* 1992. 373: 177.
- 44 Miller, J. H., *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory Press, New York 1972.
- 45 Sanger, F., Nicklen, S. and Coulson, A. R., *Proc. Natl. Acad. Sci. USA* 1977. 74: 5463.
- 46 Zachau, H. G., *Abstracts of the 2nd European Human Genome Organisation Conference*, Porto Conte, Italy, April 1992, HUGO Europe, Italy 1992, p. 11.
- 47 Maruyama, I. N., *Techniques - J. Methods Cell. Mol. Biol.* 1990. 2: 302.
- 48 Shuldiner, A. R., Nirula, A. and Roth, J., *Nucleic Acids Res.* 1989. 17: 4409.
- 49 Timmers, E., Hermans, M. M., Kraakman, M. E., Hendriks, R. W. and Schuurman, R. K., *Eur. J. Immunol.* 1993. 23: 619.
- 50 Chothia, C. and Lesk, A. M., *J. Mol. Biol.* 1987. 196: 901.
- 51 Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith, G. S., Air, G., Sheriff, S., Padlan, E. A., Davies, D., Tulip, W. R., Colman, P. M., Spinelli, S., Alzari, P. M. and Polyak, R. J., *Nature* 1989. 342: 877.

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.